Synthesis of a Genetically Engineered Repetitive Polypeptide Containing Periodic Selenomethionine Residues

Michael J. Dougherty,† Srinivas Kothakota,† Thomas L. Mason,† David A. Tirrell,‡ and Maurille J. Fournier*,†

Program in Molecular and Cellular Biology, Department of Biochemistry and Molecular Biology and Department of Polymer Science and Engineering, University of Massachusetts, Amherst, Massachusetts 01003

Received November 11, 1992
Revised Manuscript Received January 7, 1993

Genetically engineered proteins will play an important role in materials science. Many natural proteins have excellent materials properties such as the silks, elastin, and collagen, and, in principle, both these and entirely new protein materials can be produced from artificial genes. Good early progress has been made in this direction, including the synthesis of repetitive proteins predicted to self-assemble into solid lamellae of defined thickness and surface function (refs 1 and 2 and references therein). While biological synthesis offers the materials scientist superior control over polymer chain architecture, including purity of sequence, size, and stereochemistry, its versatility is limited to the 20 amino acids that appear in natural proteins. Thus, an important aim for the future will be to expand the number of amino acids that can be utilized in protein biosynthesis. First efforts in this direction would logically start with amino acid analogs known to be incorporated during translation. We describe here the first genetic synthesis of a periodic protein material in which an amino acid analog, selenomethionine, completely replaces a natural amino acid.

Early work on the amino acid specificity of translation led to the discovery that some analogs are accepted at every step in the process leading to protein synthesis. One such analog, selenomethionine, can support growth of *Escherichia coli* cells auxotrophic for methionine[3,4] and has been completely substituted for methionine in *E. coli* and phage T4 thioredoxin.[5] Other investigators have demonstrated at least partial replacement of natural amino acids with *p*-fluorophenylalanine and norleucine,[6] trifluoroleucine,[7] and other analogs.[8] The recent discovery that selenocysteine serves as the 21st naturally occurring amino acid provides further proof that translation in vivo is not limited to the 20 common amino acids.[9] However, in this case selenocysteine is not incorporated into tRNA directly; rather, serine is first charged to a suppressor tRNA and then converted to the selenium analog.[10]

A mixed chemical and biochemical strategy has been developed for incorporating backbone variants of natural amino acids into protein, providing a route for incorporating unusual structures that cannot be accomplished in vivo.[11] This technique has also been used for unnatural side-chain amino acids but requires chemical acylation, ligation to a suppressor tRNA, and in vitro translation.[12-15] As presently established, in vitro strategies are not suited to producing large quantities of analog-containing protein, and the fact that suppression efficiencies are ≤50% limits the utility of this approach in syntheses requiring amino acid replacement at multiple sites. These limitations have motivated our interest in in vivo strategies that will allow selective incorporation of novel amino acids at specified sites in repetitive protein polymers.

We have designed and expressed in vivo several protein variants consisting of repeated stem and turn domains. The stems contain a silklike sequence of alternating glycine and alanine residues, known to specify β strands.[16] Amino acids that interrupt these strands are placed at regular intervals to encourage chain reversal. Our choice of turn residues is based on two considerations. First, we reason that amino acids with polar or bulky side chains should be excluded from packing in the interior of the crystal and, therefore, will be forced to the surface. Second, the desire to produce pure materials led us to focus on amino acid analogs compatible with complete replacement. Selenomethionine satisfies both criteria and was chosen to serve as the chain defect. Glycine was also included at the turn positions to minimize restrictions on the accessible conformational space. Thus, the repetitive sequence {$(GlyAla)_3GlySeMet$} was chosen for this study.

The nucleotide sequence for this repetitive polypeptide was selected from a library of related sequences that specify variants with different two-amino-acid turns (Figure 1). The library includes codons for alanine and glycine specifying moderate to high abundance tRNAs in *E. coli*.[17-19] Since these codons are naturally GC-rich, and thereby a potential source of stably folded secondary structure, we biased the third position of the codon in favor of A and T. Eight amino acids were chosen for each stem-turn based on stereochemical considerations and molecular modeling which indicate that an even number of amino acids should place the first turn residue in a position favoring turn reversal.

DNA representing the genetic monomer was synthesized, cloned, and sequenced via the Sanger dideoxy method to verify the accuracy of the synthesis. The 24 base-pair (bp) fragment encoding one stem-turn repeat was isolated by restriction with *Ban*I, which recognizes a nonpalindromic restriction site, purified by gel electrophoresis and ligated to generate a population of multimers (data not shown). The nonpalindromic ends strongly favor the directed head-to-tail joining of the monomers. This method of assembly is critical to maintaining reading frame and sequence. The resulting population of multimers was then ligated into pMD3a, a transfer vector containing a unique *Ban*I site, and transformed into *E. coli* DH5αF′. Transformants were screened for insert size, and a sequence containing nine repeats was subcloned into the pGEX-2T expression vector at the *Bam*HI site to generate pGEX-9GM. This expression system directs the synthesis of fusion proteins containing 26 kD of glutathione *S*-transferase at the amino terminus.[20] Expression of the repetitive polymer from pGEX-9GM in *E. coli* strain HB101 generated the anticipated 31-kD fusion protein, as evidenced by Coomassie Blue R-250 staining of total protein analyzed by SDS–polyacrylamide gel electrophoresis (data not shown).

A methionine auxotroph, designated strain EM1, was prepared from *E. coli* HB101 by mutagenesis with ethyl methanesulfonate and selection with penicillin in the absence of methionine.[21] Candidate auxotrophs were screened for growth phenotype in the absence of methionine. Two isolates, EM1 and EM2, showed complete dependency on methionine. Additionally, no growth occurred in the presence of homocysteine, the penultimate product in the methionine biosynthesis pathway. Blockage at this stage of the pathway prevents methionine production from earlier intermediates. Selenomethionine, in the absence of methionine, was capable of supporting

† Department of Biochemistry and Molecular Biology.
‡ Department of Polymer Science and Engineering.

0024-9297/93/2226-1779$04.00/0

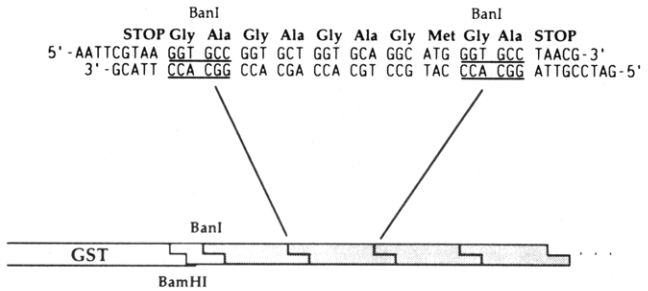

Figure 1. Nucleotide sequence of the genetic monomer and assembly into the structural gene. The monomer was cloned, sequenced, and polymerized directionally to create a repetitive coding unit. METHODS. The DNA shown was chosen from a library of sequences generated by mixed site chemical synthesis and was verified by Sanger dideoxy sequence analysis. The 24 bp fragment was isolated by restriction digestion with *Ban*I, purified, and ligated to produce a population of multimers. These were cloned into the unique *Ban*I site of the transfer vector pMD3a (not shown), and inserts were screened for size by restriction digestion at flanking *Bam*HI sites. A fragment containing nine repeats of the monomer was isolated and subcloned into the *Bam*HI site of vector pGEX-2T to yield pGEX-9GM.

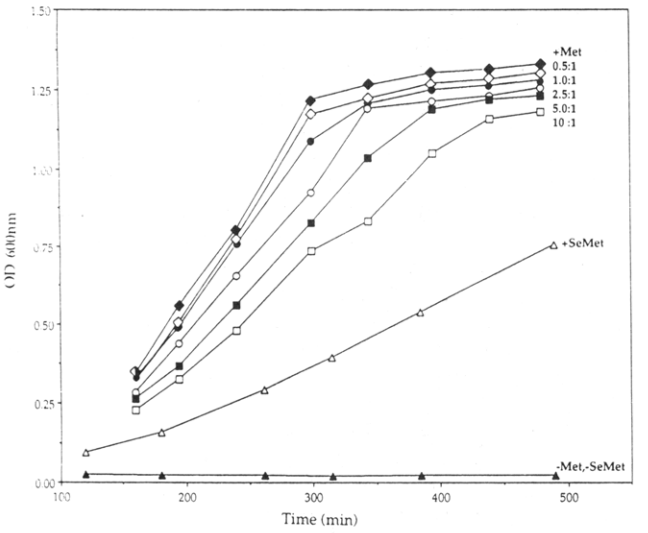

Figure 2. Growth kinetics of *E. coli* EM1/pGEX-9GM transformants grown in minimal medium containing variable amounts of selenomethionine. Curves labeled +Met, 0.5:1, 1:1, 2.5:1, 5:1, and 10:1 represent growth experiments in minimal medium supplemented with 40 μg/mL of L-methionine and either zero, 20, 40, 100, 200, or 400 μg/mL of L-selenomethionine, respectively (ratios as given are normalized [SeMet]:[Met]). The curves labeled +SeMet and −Met,−SeMet correspond to minimal-medium growth experiments containing either 20 μg/mL of L-selenomethionine or no methionine or selenomethionine, respectively. METHODS. Methionine auxotroph *E. coli* EM1 was prepared from parent strain HB101 by mutagenesis with ethyl methanesulfonate (15 μL/mL) and penicillin selection in the absence of methionine. The EM1 isolate was purified to verify its requirement for methionine. EM1 cells transformed with pGEX-9GM were grown overnight at 37 °C in 7 mL of M9 medium containing 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, 1 μg/mL of vitamin B1, 0.2% glucose, 20 μg/mL of L-methionine, and 200 μg/mL of ampicillin. Cells were collected by centrifugation (13 500*g*), washed in methionine-free medium, and resuspended in minimal medium lacking methionine. Cells were diluted 1:50 into 8 mL of medium containing variable amounts of methionine and selenomethionine (both from Sigma Chemicals). Turbidity of the cultures was monitored at 600 nm.

growth although with about a 2-fold increase in generation time (Figure 2).

The ability of selenomethionine to support growth of EM1 cells transformed with pGEX-9GM was examined at 37 °C in minimal medium supplemented with the concentrations of L-methionine and L-selenomethionine indicated in Figure 2. The growth data suggest that selenomethionine can completely replace methionine in all proteins produced in the cell.

A more direct and quantitative examination of the level of substitution in the target protein was made using a competition experiment with radiolabeled ^{35}S-methionine and unlabeled selenomethionine. Several cultures of pGEX-9GM transformants were grown to logarithmic phase in minimal medium containing 20 μg/mL of L-methionine and increasing concentrations of selenomethionine. The cultures were then radiolabeled with ^{35}S-methionine and induced 5 min later with isopropyl β-thiogalactopyranoside (IPTG). Aliquots were taken at several time points and the protein labeling patterns analyzed by SDS–polyacrylamide gel electrophoresis, Coomassie Blue staining, and fluorography. Autoradiographs and stained gels were analyzed by densitometric scanning. An approximately linear decrease in methionine incorporation was observed with increasing concentrations of unlabeled selenomethionine, consistent with complete substitution of the analog for methionine (Figure 3a). We verified that protein levels were equivalent from lane to lane by staining gels and densitometrically scanning both background and induced protein bands (data not shown).

The results shown in Figure 3b compare the scanning data obtained for the 31-kD fusion protein with the theoretical values expected if selenomethionine competes as effectively for incorporation as methionine. Experimental ratios were derived by quantifying band intensity for each time point and selenomethionine–methionine ratio and comparing these values with those obtained for the same time point but different selenomethionine–methionine ratios. The experimentally determined ratios match very closely the values expected (line in Figure 3b) if the extent of selenomethionine incorporation is determined solely by the relative concentrations of the two amino acids in the medium. The excellent correspondence of the experimental and theoretical values is consistent with complete or near-complete substitution by selenomethionine. Similar results were obtained when a prominent background *E. coli* protein was analyzed in the same way (data not shown).

As a final test of the potential to completely replace methionine in our repetitive polypeptide, an expression experiment was performed with *E. coli* EM1/pGEX-9GM transformants in medium containing selenomethionine but lacking methionine. The production of GST-9GM was monitored by separating total cell proteins on SDS–polyacrylamide gels and staining with Coomassie Blue. Upon induction the 31-kD product accumulated with a high relative yield demonstrating that this strategy can, indeed, be used to engineer selenomethionine-containing repetitive proteins (Figure 4).

Expression in this last condition is predicted to yield a repetitive protein with selenomethionines at each turn position, i.e., every eight amino acids. Folding of individual polymer chains at these defects should generate antiparallel β sheets stabilized by hydrogen bonding. The stacking of such sheets is predicted to result in a lamellar crystal with selenium atoms populating the surfaces. Structural analysis with related proteins of the class $\{(GlyAla)_nGlyX\}$, where X is an acidic amino acid, indicates that such materials do form ordered β sheets in the solid state (manuscripts in preparation). In such a configuration the exposed selenides would be available for post-translational modification, for example, oxidation to form reactive selenoxides and alkylation to form trialkylselenonium

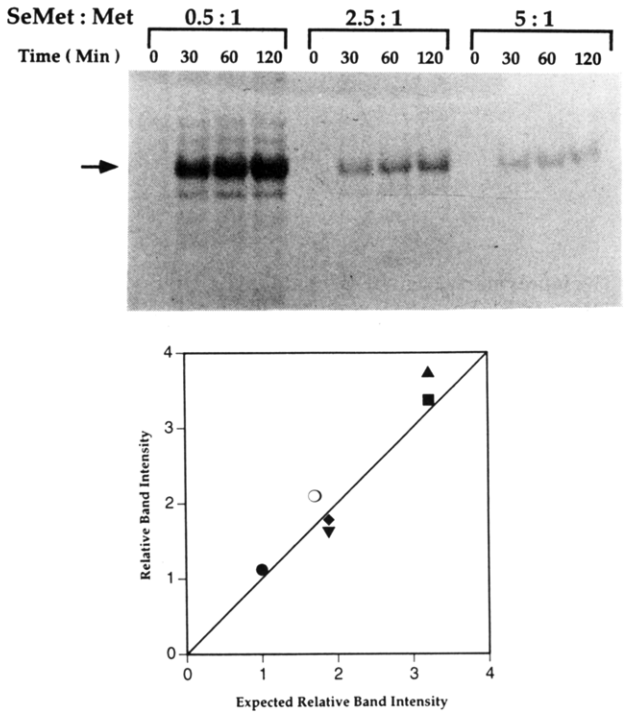

Figure 3. Effect of selenomethionine on the incorporation of ^{35}S-methionine into the repetitive polypeptide GST-9GM. (a) Autoradiograph of 12.5% SDS–polyacrylamide gel showing total cell protein. Each set of lanes represents a different ratio of selenomethionine to methionone. Cultures were induced at t = 0. The position of the target protein is indicated by an arrow. (b) Comparison of experimental and predicted incorporation values for the induced polypeptide at different ratios of selenomethionine to methionine. Experimental ratios were obtained by scanning autoradiographic signals. The theoretical ratios correspond to the intensities expected if the incorporation of selenomethionine is equivalent to that of methionine at each concentration tested. METHODS. (a) Transformants were grown in 5 mL of minimal medium at the same selenomethionine–methionine ratios described in Figure 2. Cultures were labeled wih 25 μCi ^{35}S-methionine (23.4 μCi/μmol) at OD(600 nm) = 0.4. Five minutes later (t = 0) the cells were induced with 0.4 mM isopropyl β-galactopyranoside. Aliquots were harvested at time points 0, 30, 60, and 90 min, washed once with minimal medium, and lysed in sample buffer (0.06 M Tris-Cl, pH 6.8, 2% SDS, 10% glycerol, 0.025% bromophenol blue, and 5% 2-mercaptoethanol). Cell extract corresponding to equal numbers of cells was loaded in each lane, and the proteins were separated by SDS–polyacrylamide gel electrophoresis. Direct labeling experiments with radiolabeled ^{75}Se-methionine were not possible due to difficulties in finding reliable suppliers. (b) Gels and autoradiographs were scanned using a MicroScan 1000 two-dimensional gel analyzer (Technology Resources, Inc., Nashville, TN). Autoradiographic signals corresponding to the induced protein were compared in a pairwise fashion at identical induction times for several different selenomethionine–methionine ratios. 10X Met represents a control consisting of a 10-fold excess of unlabeled methionine. Each experimental value consists of intensity measurements taken from at least three separate time points and averaged to yield the experimental ratio. ●, 10:1/10X Met; ○, 2.5:1/5:1; ◆, 5:1/10X Met; ▼, 5:1/10:1; ■, 2.5:1/10:1; ▲, 2.5:1/10X Met. Theoretical ratios are represented by the line.

halides.[22] Analogous reactions with sulfur methionine produce less reactive intermediates.

These results demonstrate that an unnatural amino acid can be utilized in biological synthesis to yield a novel polypeptide material. Since this material possesses a chemical function beyond those provided by the natural amino acids, it is now feasible to produce "designer" protein materials while still retaining the advantages of in vivo

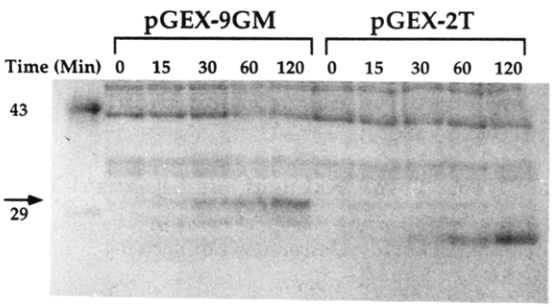

Figure 4. Induction of recombinant protein in medium containing selenomethionine but no methionine. Proteins were detected by staining an SDS–polyacrylamide gel with Coomassie Blue. METHODS. Transformed *E. coli* EM1 cells were grown overnight in M9 minimal medium supplemented with 20 μg/mL of L-methionine and 200 μg/mL of ampicillin. Cells were pelleted, washed twice in minimal medium lacking methionine, and inoculated into minimal medium containing L-selenomethionine at 20 μg/mL. Expression was induced 6 h later with 0.4 mM IPTG. As a control, the glutathione *S*-transferase protein from pGEX-2T was induced under identical conditions in HB101. Size standards (kD) are present in the first lane. The target protein is identified with an arrow.

production. This finding provides optimism that many other analogs known to be utilized by *E. coli* can be incorporated into genetically engineered protein materials. In principle, it should be feasible to obtain complete substitution of analogs that do not support growth by using inducible expression systems. While this would be an exciting and important development, even partial replacement will be valuable for many applications.

Acknowledgment. This work was supported by a grant from the Polymer and Genetics programs of the National Science Foundation (DMR 8914359).

References and Notes

(1) McGrath, K. P.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. *J. Am. Chem. Soc.* **1992**, *114*, 727.
(2) Tirrell, D. A.; Fournier, M. J.; Mason, T. L. *Curr. Opin. Struct. Biol.* **1991**, *1*, 638.
(3) Tuve, T.; Williams, H. *J. Am. Chem. Soc.* **1957**, *79*, 5830.
(4) Cowie, D. B.; Cohen, G. N. *Biochim. Biophys. Acta* **1957**, *26*, 252.
(5) Hendrickson, W. A.; Horton, J. R.; LeMaster, D. M. *EMBO J.* **1990**, *9*, 1665.
(6) Cowie, D. B.; Cohen, G. N.; Bolton, E. T.; De Robichon-Szulmajster, H. *Biochim. Biophys. Acta* **1959**, *34*, 39.
(7) Fenster, E. D.; Anker, H. S. *Biochemistry* **1969**, *8*, 269.
(8) Richmond, M. H. *J. Mol. Biol.* **1963**, *6*, 284.
(9) Zinoni, F.; Birkmann, A.; Stadtman, T. C.; Bock, A. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 4650.
(10) Leinfelder, W.; Zehelein, E.; Mandrand-Berthelot, M. A.; Bock, A. *Nature* **1988**, *331*, 723.
(11) Ellman, J. A.; Mendel, D.; Schultz, P. G. *Science* **1992**, *255*, 197.
(12) Mendel, D.; et al. *Science* **1992**, *256*, 1798.
(13) Noren, C. J.; Anthony-Cahill, S. J.; Griffith, M. C.; Shultz, P. C. *Science* **1989**, *244*, 182.
(14) Heckler, T. G.; et al. *Tetrahedron* **1984**, *40*, 87.
(15) Heckler, T. G.; et al. *Biochemistry* **1984**, *23*, 1468.
(16) Fraser, R. D. B.; MacRae, T. P.; Stewart, F. H. C. *J. Mol. Biol.* **1966**, *19*, 580.
(17) Ikemura, T. *Mol. Biol. Evol.* **1985**, *2*(1), 13.
(18) Andersson, S. G. E.; Kurland, G. C. *Microbiol. Rev.* **1990**, *54*, 198.
(19) Aota, S.; et al. *Nucleic Acids Res.* **1988**, *16*, r315.
(20) Smith, D. B.; Johnson, K. S. *Gene* **1988**, *67*, 31.
(21) Miller, J. H. *Experiments in Molecular Genetics*; Cold Spring Harbor Laboratory: Cold Spring Harbor, NY, 1972.
(22) Paulmier, C. *Selenium Reagents and Intermediates in Organic Synthesis*; Pergamon Press: New York, 1986.